

Enhanced real-time glaucoma diagnosis: dual deep learning approach

Mai Hesham¹, Ghada Kareem², Marwa Hadhoud¹

¹Department of Biomedical Engineering, Faculty of Engineering, Helwan University, Cairo, Egypt

²Department of Biomedical Engineering, Higher Technological Institute, Sharkeya, Egypt

Article Info

Article history:

Received Mar 22, 2024

Revised Nov 27, 2024

Accepted Dec 24, 2024

Keywords:

Deep learning

Glaucoma detection

Optic cup segmentation

Optic disc segmentation

U-Net

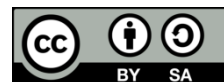
You only look at coefficients

You only look once version 5

ABSTRACT

Effective management of glaucoma is essential for preventing irreversible vision loss. This study introduces a novel deep learning-based network designed to enhance performance while minimizing computational complexity. The system comprises two models: the first is a hybrid model combining a customized U-Net architecture integrated with you only look at coefficients (YOLACT) is utilized to achieve accurate segmentation of the optic disc (OD) and optic cup (OC), providing detailed diagnostic insights for ophthalmologists. The second model employs you only look once version 5 (YOLOv5) for real-time glaucoma prediction, delivering outstanding performance with an accuracy of 97.89% and F1 score of 98% on the primary dataset. On an independent dataset without further training, the model achieved 96% accuracy, with sensitivity and specificity of 98.9% and 93.3%, respectively. These results highlight the model's robustness, generalizability, and adaptability, demonstrating its potential for effective glaucoma screening and early detection in diverse clinical environments. This approach offers a promising advancement in improving the accessibility and efficiency of glaucoma management.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Mai Hesham

Department of Biomedical Engineering, Faculty of Engineering, Helwan University

Cairo, Egypt

Email: mai.hesham.abdelkader@h-eng.helwan.edu.eg

1. INTRODUCTION

Glaucoma, a prevalent ophthalmic condition, leads to progressive damage to the optic disc (OD) and optic nerve head (ONH), crucial structures linking the eye to the brain, which results in gradual deterioration of the visual field [1], [2]. This condition is often described as having asymptomatic early stages, with symptoms becoming evident only as the disease advances, potentially leading to severe peripheral vision loss or total blindness [3]. By 2040, the number of glaucoma cases worldwide is projected to reach 111.8 million [4], making it a leading cause of irreversible blindness and contributing to 20% of global blindness cases [5]. Glaucoma affects approximately 0.5% of individuals under fifty and up to 10% of those over eighty [6]. Early detection and proper treatment are vital to prevent these serious consequences [7].

The diagnosis of glaucoma depends on detecting optic nerve cupping and changes in the central portion of the OD. Examining a fundus image of a patient's eye manually requires the expertise of specialized ophthalmologists [8]. While expert ophthalmologists can interpret OD images to some extent, machine learning classifiers often outperform human observers [9]. Accurate glaucoma detection hinges on the precise segmentation of the OD and optic cup (OC), which presents significant challenges, primarily owing to the obstructive influence of blood vessels.

Various approaches for OD and OC segmentation have been explored, including both deep learning and non-deep learning methods. While non-deep learning methods are typically more efficient in terms of computational speed, they often struggle to achieve the same level of accuracy as deep learning techniques. The following sections provide a detailed review of these methods and discuss how they inform the current approach, beginning with non-deep learning approaches and followed by recent advancements in deep learning techniques, each of which offers unique advantages and limitations for OD and OC segmentation.

Several studies have employed non-deep learning approaches to achieve OD and OC segmentation, often utilizing similar techniques to enhance performance. Issac *et al.* [6], Zainudin *et al.* [10] both utilized adaptive thresholding as part of their segmentation processes. Issac *et al.* [6] included features like mean and standard deviation while extracting pertinent data from the red and green channels in fundus images, reaching an accuracy of 92.06%. Zainudin *et al.* [10], in contrast, applied color channel separation for pre-processing to eliminate noise from fundus images before using adaptive thresholding and morphological processing for segmentation, resulting in a peak signal-to-noise ratio (PSNR) value of 64.09. Similarly, Nugroho *et al.* [9], Septiarini *et al.* [11] employed morphological methods to segment the OC and detect glaucoma. Nugroho *et al.* [9] combined morphological reconstruction with convex hull techniques for OC segmentation and used active contouring for OD, reaching accuracies of 96.3% and 99.13% for OD and OC segmentation, respectively. However, the study was limited by the small size of the dataset used. Septiarini *et al.* [11], on the other hand, introduced cup area contour analysis using morphological methods and thresholding to extract ONH features, achieving a 94.44% accuracy with a support vector machine (SVM) classifier. For feature selection, Singh *et al.* [12], [13] explored optimization algorithms. Singh *et al.* [12] proposed the gravitational search optimization algorithm (GSOA) to identify influential features and trained six machine learning models, reaching an accuracy of 95.36%. In a subsequent study, Singh *et al.* [13] developed a hybrid algorithm by merging emperor penguin optimization (EPO) with bacterial foraging optimization (BFO), which, when integrated with a random forest classifier, achieved a peak accuracy of 95.41%. Nevertheless, non-deep learning methods that depend on features like texture, color, and gradient tend to be less effective when dealing with low-contrast images.

In contrast, several studies have employed deep learning networks for OD and OC segmentation, as well as for the direct classification of fundus images, often sharing common techniques and approaches. Civit-Masot *et al.* [14] utilized U-Net-based models in a dual-subsystem framework for OD and OC segmentation. They combined this with a MobileNet V2 network for direct classification of fundus images, while in another study Civit-Masot *et al.* [15] employed a Keras U-Net model known for its generalizability and configurability, however memory constraints posed challenges during training. Similarly, Sevastopolsky [16] applied a modified U-Net for automatic OD and OC segmentation, incorporating an image enhancement technique was applied through a preprocessing step using contrast limited adaptive histogram equalization (CLAHE) to improve quality of images.

Other researchers, such as Al-Bander *et al.* [3], Zilly *et al.* [17] introduced further modifications to CNN-based models to improve OD and OC segmentation. Zilly *et al.* [17] proposed a novel convolutional neural networks (CNN) method that utilized an entropy-based sampling technique to reduce computational complexity, offering superior results compared to uniform sampling, though the method's complexity poses challenges for reproducibility. Al-Bander *et al.* [3] developed a heavily modified dense U-Net, evaluating its performance across five datasets. However, both Al-Bander *et al.* [3], Zilly *et al.* [17] approaches rely on cropping images around the OD/OC area prior to segmentation, making them unsuitable for processing full, unseen fundus images.

Singh *et al.* [18] extended the application of advanced deep learning networks, using multiple deep learning methods for identifying the retinal blood vessel network in the STARE dataset. Among these models, LadderNet achieved the highest accuracy of 97.1%, highlighting the effectiveness of multi-stage, ladder-style architectures in segmentation tasks. Afolabi *et al.* [19] also modified the U-Net for OD and OC segmentation but combined it with an extreme gradient boosting (XGB) algorithm for glaucoma detection.

Ramesh *et al.* [20], Anusha *et al.* [21] employed custom you only look once version 5 (YOLOv5) architectures for glaucoma detection. Ramesh *et al.* [20] used CNN and an approach that incorporates human input for data labeling. However, obtaining high accuracy with lower-resolution fundus images remains challenging. while Anusha *et al.* [21] trained YOLOv5 on the ORIGA dataset and tested it on the REMEDIO dataset to avoid overfitting. A summary of the previously mentioned studies is presented in Table 1.

A common limitation of the previously discussed models is the complexity required to achieve high performance, along with the need to crop images to a region of interest (RoI) before training. To overcome these challenges, the proposed system introduces an efficient and high-speed network composed of two deep learning models designed as a diagnostic tool for glaucoma detection. The first model employs a hybrid approach that integrates U-Net and you only look at coefficients (YOLACT), leveraging U-Net's strong spatial detail preservation and YOLACT's efficient instance segmentation at multiple scales for precise segmentation. The second model uses YOLOv5s, which features a low parameter count and a notably fast

runtime, to detect glaucoma by analyzing full fundus images of the eye. Moreover, the YOLOv5s model was evaluated on the PAPILA dataset without further training, demonstrating the effectiveness of transfer learning and enhancing the model's applicability in glaucoma detection. The suggested system not only delivers cutting-edge accuracy but also provides a valuable tool for monitoring structural changes in both the OC and OD. This work aims to deliver an efficient, fast, and accurate system to help doctors detect glaucoma early, a critical step in preventing permanent vision impairment. The upcoming sections will detail the method used in the proposed system (section 2), present and discuss the results obtained (section 3), and provide conclusions drawn from this study (section 4).

Table 1. Overview of previous research

Ref	Dataset	Model	Aim	Evaluation metrics
[3]	RIM-One (r3), DRISHTI, and DRIONS	Dense U-Net	OD/OC segmentation	RIM-ONE (r3): OD Dice=90%, DRIONS: OD Dice=94%, DRISHTI: OD Dice=95%
[9]	50 images from the DRISHTI	-		OD Acc.=96.3%, OC Acc.=99.13%
[14]	RIM-ONE (r3) and DRISHTI	Generalized U-Net		RIM-ONE (r3): OD Dice=92%, OC Dice=84%, DRISHTI: OD Dice=93% OC Dice=89%
[15]	RIM-One (r3), DRISHTI, and DRIONS	Generalized U-Net		Multi-dataset: Dice OC= Dice OD=94%
[16]	RIM-One (r3), DRISHTI, and DRIONS	Modified U-Net		RIM-ONE (r3): Dice OD=95%, OC Dice=82%, DRIONS: Dice OD=94%, DRISHTI: OC Dice=85%
[17]	DRISHTI	CNN		OC Dice= 83%
[19]	RIM-One (r3), DRISHTI, and DRIONS	U-Net Lite		RIM-ONE (r3): Dice= 96%, DRIONS: OD Dice=97%, DRISHTI: OD Dice=97%
[6]	63 images from the local hospital	-	Glaucoma classification	Acc.=92.06%
[12]	3112 images (ACRIMA-DRISHTI-HRF-ORIGA and private dataset)	RF		Acc.=95.36%
[13]	3112 images (ACRIMA-DRISHTI-HRF-ORIGA and private dataset)	RF		Acc.=95.41%
[10]	RIM-ONE (r3), DRIONS	-		PSNR=64.09
[11]	44 images from Dr. YAP Eye Hospital	SVM		Acc.=94.44%
[18]	STARE	LadderNet		Acc.=97.1%
[20]	Private	YOLOv5		Acc.=98%
[21]	ORIGA	YOLOv5		Acc.=92%

2. METHOD

The complete system that was developed and trained is visually summarized in Figure 1 for the first model and Figure 2 for the second model. These figures offer a detailed summary of the training and evaluation procedures conducted for each model.

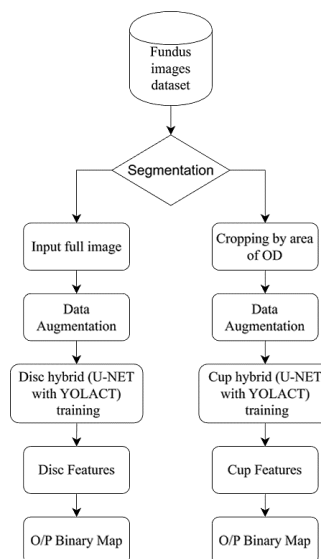


Figure 1. First model, disc, and cup segmentation using hybrid U-Net model

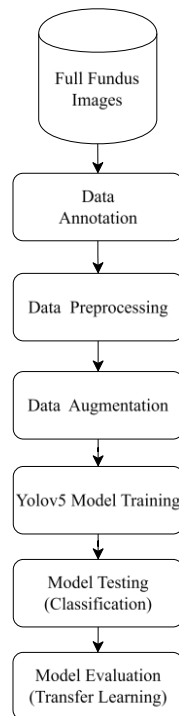


Figure 2. Second model, direct glaucoma detection using YOLOv5

2.1. Dataset

The segmentation model utilized publicly available three different datasets, namely RIMONE (r3), DRISHTI, and DRIONS, in our study. The RIMONE (r3) dataset, provided by the MIAG group at the University of La Laguna (Spain), comprises 159 fundus images that have been cropped based on the OD area. These images were expertly labeled by ophthalmologists for both the OD and OC [22]. The DRISHTI-GS dataset, obtained from Aravind Eye Hospital in Madurai (India), consists of 101 full fundus images that have also been labeled for the OD and OC [23]. Finally, the DRIONS-DB dataset, sourced from Miguel Servet Hospital in Saragossa (Spain), contains 110 full eye fundus images [24].

The direct classification model trained by a combined dataset sourced from various resources, and we aim to demonstrate its generalizability and applicability to real-world data. Specifically, we trained our model on the previously illustrated DRISHTI-GS dataset combined with the publicly available ORIGA dataset [23], [25], which consists of 650 fundus images. The ORIGA dataset included 482 normal eyes and 168 glaucomatous eyes. We used only 168 of the healthy images for data imbalance reduction.

2.2. The segmentation model

The first model utilizes a two-stage hybrid U-Net model for segmenting OD and OC. This section describes the steps implemented for the segmentation of the OD and OC.

2.2.1. Pre-processing

To enable the utilization of dataset images in the segmentation model, the following steps are essential:

- For OC segmentation, the images need to undergo cropping based on the bounding box of the OD.
- For OD segmentation the full image is resized to 256×256 pixels, whereas the OC segmentation focuses on the cropped area that is resized to 128×128 pixels for processing efficiency and accuracy.

2.2.2. Data augmentation

During the training of our system, static techniques for data augmentation were used to expand the training set by generating synthetic samples. This process involved rotation at an angle of 50°, a height shift of 0.15, a fill mode (with a constant value of 0), and horizontal and vertical flipping, as shown in Figure 3.

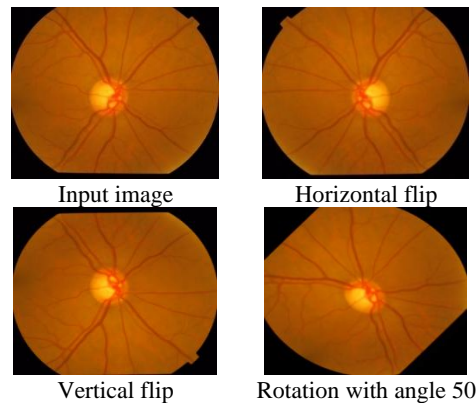


Figure 3. Visual example of static data augmentation applied to fundus images

2.2.3. System architecture

In this study, we leverage a modified U-Net architecture, originally proposed by Sevastopolsky [16], to address the segmentation complexities associated with OC and OD. The presented model is a hybrid architecture that combines elements of the U-Net and YOLACT frameworks to create a robust approach for image segmentation. The U-Net represents a fully convolutional deep learning network specifically crafted for semantic segmentation tasks within the realms of image processing and computer vision. Its efficacy has been demonstrated in numerous medical segmentation challenges. While YOLACT is developed for real-time instance segmentation in the field of computer vision [26]. Provided below, the detailed description of the architecture, a graphical representation is provided in the supplementary material.

a. U-Net components:

- Encoder: the initial layers of the model act as the encoder. These layers use multiple convolutional blocks and residual blocks to progressively downsample the input image, capturing both low-level and high-level features. The encoded feature maps at different resolutions are stored and passed to later layers.
- Decoder: the feature pyramid network (FPN) in the model acts like a U-Net decoder. The FPN upsamples lower-resolution feature maps to match the resolution of higher-resolution feature maps. The feature maps from the encoder are used as lateral connections, which are combined with the upsampled feature maps. This mirrors the skip connections in U-Net and helps the model retain high-level semantic information while also refining details through shallow feature maps (which contain more fine spatial details).

b. YOLACT components:

- Backbone structure: the initial part of the network (the encoder) performs feature extraction in a manner similar to YOLACT's backbone. This part of the model uses a combination of convolutional and residual blocks to extract feature maps at different scales. The residual blocks help this model maintain the "deep feature extraction" capability typical of YOLACT's backbone structure.
- FPN: YOLACT uses an FPN to generate feature maps at different scales, which is a key component for instance segmentation. This hybrid model implements an FPN to upsample and merge feature maps from different layers of the encoder. This allows for multi-scale feature aggregation, similar to YOLACT's approach to combining features from different resolutions.
- Mask heads: YOLACT uses separate mask heads to generate segmentation masks for different instances. In this hybrid model a series of convolutional layers are applied to generate masks for each class.
- Mask fusion: after the mask heads generate masks at different scales, these are upsampled and concatenated along the channel axis. This resembles YOLACT's mask coefficient-based approach where different mask predictions are combined to form the final output.

2.2.4. Training

During the training phase, the network is instantiated with a learning rate of 10^{-4} . Optimization was carried out using Adam optimizer. Our experiments run on a Tesla T4 GPU in Google Colab.

2.2.5. Experiments

We conducted a series of experiments on eye fundus images to perform segmentation of both the OD and OC. For each dataset, we partitioned the data into separate training and testing sets, using 80% for training and 20% for testing. For OD segmentation, the experiments utilized the DRISHTI, DRIONS-DB,

and RIM-ONE (r3) datasets. It is important to note that for the DRIONS-DB dataset, results are only provided for OD segmentation, as the ground truth for OC is unavailable [3], [19]. For OC segmentation, we focused on the DRISHTI and RIM-ONE (r3) datasets. To evaluate the performance of the segmentation model, we used the intersection over union (IoU) and dice score metrics, which are standard in image segmentation and computer vision tasks. The highest performance of our model for OD segmentation on the RIM-ONE (r3) dataset is illustrated in Figure 4, while the best results for OD segmentation on the DRIONS-DB dataset are shown in Figure 5.

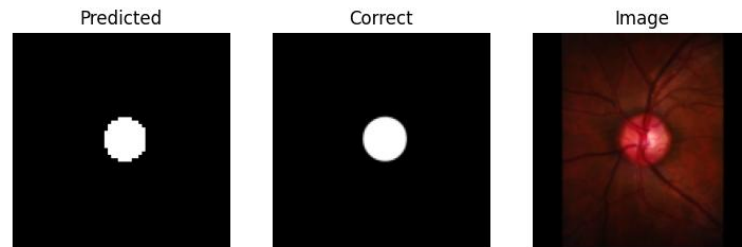


Figure 4. A visual comparison between the predicted outcomes and the correct segmentations on RIM-ONE (r3) for OD segmentation: best case (IoU=0.93, Dice=0.95)

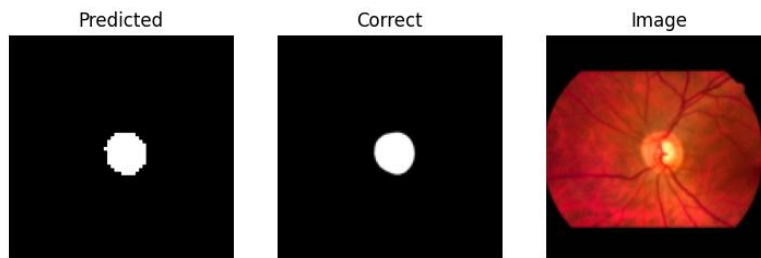


Figure 5. A visual comparison between the predicted outcomes and the correct segmentation on DRIONS-DB for OD segmentation: best case (IoU=0.93, Dice=0.95)

2.3. Direct detection model

The second model entails the utilization of the YOLOv5s for the direct classification of fundus images without an intermediary segmentation process. In this approach, entire images are employed for training the YOLOv5s model. Upon an extensive review of research papers authored by various scholars, we discovered that the YOLOv5 algorithm is capable of efficiently diagnosing and classifying glaucomatous fundus images [27]. These networks typically exhibit the ability to recognize diverse patterns across different images and objects using the same architecture. YOLOv5 dissects the fundus into numerous segments rather than treating the entire fundus image as a singular entity [20]. This section describes the steps implemented for this model.

2.3.1. Data annotation

The foundation of any successful artificial intelligence (AI) algorithm lies in data annotation. A machine learning algorithm capable of image detection without annotated data is implausible. Human interaction becomes pivotal in the data annotation process to effectively address the black box dilemma [20]. The dataset was uploaded to the Roboflow Annotate website. These images are manually annotated using the Roboflow Annotate annotation tool [28], with reference to the ground truth (RoI). This involves drawing a loose bounding box around the OD to quantify accuracy. The annotated data were then formatted in YOLO (.txt files). Each image is paired with a corresponding .txt file of the same name, which contains annotated bounding boxes for the target object, the OD. The annotation file includes details such as the object class, coordinates, height, and width of the ground truth object within the image [29]. Utilizing Roboflow offered a significant benefit owing to its user-friendly interface and smart default settings. This streamlined the image annotation process, enabling us to annotate images quickly and precisely [30].

2.3.2. Pre-processing

The datasets were divided into 60% for training and 20% for validation, and the remaining 20% of the data were reserved as an independent test set. The images were resized to (640×640) pixels using Roboflow to fit into the 15 gigabytes GPU memory [28].

2.3.3. Data augmentation

To augment the images, we employed the "Roboflow" platform, which offers various data augmentation techniques. These techniques included flipping (horizontal and vertical) to make the model insensitive to subject orientation and to facilitate image interpretation from various perspectives and viewpoints, adjusting brightness to enhance the model's resilience to changes in lighting and camera settings, and modifying exposure. The modification of exposure is essential due to the presence of diverse lighting conditions. This approach plays a crucial role in enabling the model to generalize and learn to perform tasks effectively across a wide range of lighting scenarios [30].

2.3.4. You only look once version 5 architecture

YOLOv5 stands as a real-time object detector, building upon the advancements of YOLOv1 to YOLOv4. The network architecture of YOLOv5 is visually depicted in Figure 6 [31]. The network is divided into three key sections: the backbone, the neck, and the head [31].

- Backbone: YOLOv5 incorporates the cross stage partial network (CSPNet) into DarkNet, forming CSPDarkNet as its backbone for feature extraction from images. By utilizing residual and dense blocks, CSPDarkNet facilitates information flow to the deepest layers, mitigating the vanishing gradient problem. in contrast to other large-scale CNN, CSPNet addresses issues related to repeated gradient information, integrating gradient changes into the feature map.
- Neck: by utilizing a path aggregation network (PANet), a FPN is generated to aggregate features, which are subsequently passed to the head for prediction.
- Detection head: the head of YOLOv5 consists of three convolution layers predicting the location of bounding boxes (x, y, height, and width), scores, and object classes.

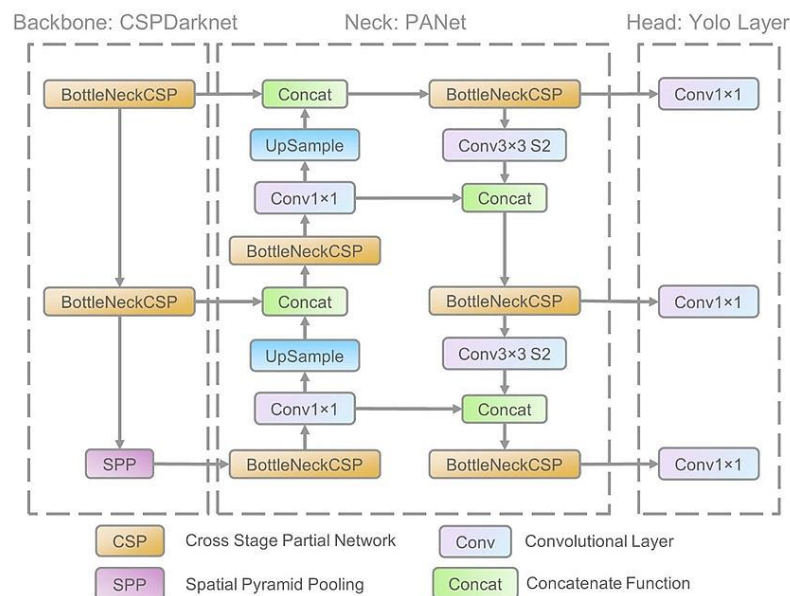


Figure 6. The network architecture of YOLOv5

2.3.5. Training

Following the completion of labeling and preprocessing, the dataset was exported to YOLOv5 PyTorch and seamlessly integrated into the code within Google Colab for further analysis and model training. Following the dataset import, the training process starts using YOLOv5s, with the "train.py" file in the YOLOv5s directory employed for model training. Figure 7 shows a sample of batch data predictions during training of YOLOv5s, encompassing class probabilities, objectness scores, and bounding boxes that apply the anchor boxes on normal and glaucomatous fundus features.

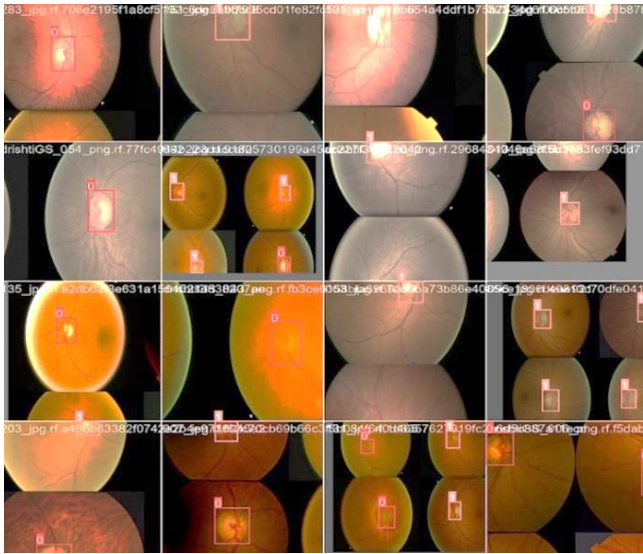


Figure 7. A sample of batch data predictions during training

3. RESULTS AND DISCUSSION

This section presents and analyzes the results obtained from our proposed system, offering a detailed comparative assessment with existing methods. The analysis focuses on both segmentation and direct detection models, highlighting the strengths and limitations of our approach relative to the current state-of-the-art.

3.1. The segmentation model

In this section, we present a comparative evaluation of our OD and OC segmentation results with those from other studies that also use deep learning-based segmentation methods on the same fundus image datasets. The performance of our model for OD and OC segmentation is summarized in Tables 2 and 3, respectively. The results from our experiments show that the proposed approach delivers competitive performance across multiple evaluation metrics while maintaining a significantly lower parameter count (14×10^6) compared to the original U-Net model's (3.1×10^7) parameters [19]. Moreover, the model demonstrates exceptional speed and efficiency in real-time instance segmentation, achieving a runtime of 0.027 seconds for OD segmentation and 0.022 seconds for OC segmentation.

Table 2. Comparison of methods for OD segmentation

Author	Method	DRIONS-DB		RIM-ONE (r3)		DRISHTI	
		IoU	Dice	IoU	Dice	IoU	Dice
Sevastopolsky [16]	Modified U-Net	0.89	0.94	0.89	0.95	-	-
Civit-Masot <i>et al.</i> [14]	Generalized U-Net	-	-	-	0.92	-	0.93
Al-Bander <i>et al.</i> [3]	DenseNet	-	0.94	0.83	0.90	0.90	0.95
Afolabi <i>et al.</i> [19]	U-Net Lite	0.90	0.96	0.90	0.96	0.90	0.97
Proposed approach	Hybrid U-Net	0.88	0.93	0.88	0.93	0.94	0.96

Table 3. Comparison of methods for OC segmentation

Author	Method	RIM-ONE (r3)		DRISHTI	
		IoU	Dice	IoU	Dice
Sevastopolsky [16]	Modified U-Net	0.69	0.82	0.75	0.85
Civit-Masot <i>et al.</i> [14]	Generalized U-Net	-	0.84	-	0.89
Al-Bander <i>et al.</i> [3]	DenseNet	0.56	0.69	0.71	0.83
Afolabi <i>et al.</i> [19]	U-Net Lite	0.73	0.86	0.79	0.95
Zilly <i>et al.</i> [17]	CNN	-	-	0.86	0.83
Proposed approach	Hybrid U-Net	0.73	0.84	0.86	0.88

3.2. Direct detection model

The effectiveness of the proposed system is evaluated in two distinct scenarios. In the first scenario, the model is trained and tested on a merged dataset consisting of ORIGA and DRISHTI images. In the

second, the model is evaluated on the independent PAPILA dataset, without any further training. To assess the performance of the YOLOv5s model, we use standard evaluation metrics, including accuracy, precision, recall, and F1-score, which are commonly applied in object detection and classification tasks [28].

This section offers an in-depth comparison of our approach with current methods for glaucoma detection. The results of our proposed technique for direct glaucoma detection are presented in Table 4, where we compare its performance with existing approaches in the literature based on metrics such as accuracy, F1-score, and run time. Table 4 presents a comparison of the models based on their accuracy. Our model achieved an impressive accuracy of 97.89%, with precision and recall (sensitivity) both reaching 98%, outperforming the results of [14], which reported an accuracy of 86% and a sensitivity of 89%. Additionally, our model attained an F1 score of 98%, surpassing the performance of [11], [21]. A crucial advantage of our approach is its speed, which is vital for real-world applications that demand quick processing. The proposed model demonstrates an outstanding average runtime of 0.02 seconds per image. Additionally, it boasts a relatively small number of parameters, totaling (7.6×10^6) . The YOLOv5s model excels in detection speed and response time, essential for the early detection of glaucoma, underscoring its efficiency in real-world applications.

Table 4. A comparison of the proposed method for direct glaucoma detection with existing techniques in the literature

Author	Method	Dataset	Accuracy	F1 score
Septiaini <i>et al.</i> [11]	SVM	44 images from Dr. YAP Eye Hospital	0.94	0.94
Anusha <i>et al.</i> [21]	YOLOv5	ORIGA	0.92	0.91
Ramesh <i>et al.</i> [20]	YOLOv5	Private dataset	0.98	-
Civit-Masot <i>et al.</i> [14]	MobileNet V2	RIM-ONE (r3)-DRISHTI	0.86	-
Proposed approach	YOLOv5s	ORIGA-DRISHTI	0.97	0.98

3.3. Model evaluation

To ensure generalizability we further tested our YOLOv5s model on another dataset without further training, utilizing the publicly available PAPILA dataset [32], comprising fundus images from both eyes of the same patient. The dataset is categorized into three classes: glaucomatous, non-glaucomatous, and suspect. The dataset comprises 488 JPEG images corresponding to the right and left eyes of 244 patients. Notably, this dataset represents a more challenging scenario because it involves a classification problem involving three classes. The "suspect" class poses an undetermined clinical diagnostic, creating complexity in the classification decision [33]. The suspect class exhibited characteristics that did not align with being a borderline class between the healthy and glaucoma classes. Instead, it behaves as an unresolved class positioned between the two, posing a challenge in finding a clear separation boundary between these classes. Unlike the method of [32], which excluded the suspect class from the original dataset, our proposed method utilized the entire dataset without removing any images by combining the glaucomatous and suspect class as one class to apply binary classification. Our system achieved an accuracy of 96% with sensitivity and specificity of 98.9% and 93.3%, respectively.

3.4. Discussion

This study aimed to advance the field of glaucoma detection through improved segmentation of the OD and OC and enhanced classification accuracy for glaucoma. Previous research has explored various approaches, including both non-deep learning and deep learning methods, for OD and OC segmentation and glaucoma detection. However, many existing methods either fall short in terms of accuracy, use complex models to achieve higher accuracy, or require RoI cropping which limits their applicability to new, unseen images.

Our approach, utilizing a modified U-Net for OD and OC segmentation and YOLOv5 for glaucoma detection, demonstrates robust performance across several datasets. Specifically, the modified U-Net which is a hybrid model that combines U-Net's strength (high spatial resolution), the model retains the U-Net's ability to preserve spatial resolution by using skip connections between corresponding encoder and decoder layers through the FPN. This helps capture fine details in the image that are important for pixel-wise segmentation tasks. YOLACT's strength (real-time instance segmentation), by integrating YOLACT's feature extraction and mask head ideas, the model can efficiently generate instance masks at multiple scales. The combination of multiple-scale feature maps through the FPN and mask heads allows the model to handle objects of varying sizes, a hallmark of YOLACT.

Additionally, the second model utilizes YOLOv5s which attained without additional training an accuracy of 96%, with a sensitivity of 98.9% and specificity of 93.3% on the PAPILA dataset, demonstrating high precision, generalizability and adaptability. Our findings are consistent with and extend the results of previous studies. The high accuracy and F1 score achieved by the YOLOv5 model align with recent advancements in deep learning for glaucoma detection, reflecting improvements over older methods that

faced challenges with image resolution and classifier performance. The proposed system is both effective and reliable for glaucoma detection and can serve as a valuable second opinion tool for medical practitioners. Additionally, due to its high performance, the proposed system can be utilized to create mobile applications that assist physicians in the early detection of glaucoma. Consequently, the rate of vision loss can be reduced.

4. CONCLUSION

This study presents an innovative deep learning-based approach for OD and OC segmentation, as well as glaucoma detection, which surpasses existing methods in both accuracy and computational efficiency. The modified U-Net model and the YOLOv5-based detection framework exhibit high effectiveness, making them well-suited for real-time clinical applications, especially in resource-limited environments. The models' capability to maintain high performance across diverse datasets, including complex cases, underscores their potential for enhancing early glaucoma diagnosis. These advancements suggest a promising trajectory for the development of automated tools that could significantly impact clinical practice by enabling faster, more accurate glaucoma detection and potentially improving patient outcomes. Future research could explore alternative deep learning architectures or hybridization techniques that may further enhance accuracy and efficiency. Additionally, evaluating these models on larger, more diverse datasets could provide deeper insights into their generalizability in real-world clinical settings.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mai Hesham	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ghada Kareem	✓	✓		✓	✓	✓				✓	✓	✓	✓	
Marwa Hadhoud	✓	✓		✓	✓	✓				✓	✓	✓	✓	✓

C : **C**onceptualization

M : **M**ethodology

So : **S**oftware

Va : **V**alidation

Fo : **F**ormal analysis

I : **I**nterpretation

R : **R**esources

D : **D**ata Curation

O : **O**riginal Draft

E : **E**diting

Vi : **V**isualization

Su : **S**upervision

P : **P**roject administration

Fu : **F**unding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY





Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES





- [1] R. Ali *et al.*, "Optic disk and cup segmentation through fuzzy broad learning system for glaucoma screening," in *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2476-2487, Apr. 2021, doi: 10.1109/TII.2020.300020.
- [2] L. K. Singh, M. Khanna, and R. Singh, "Feature subset selection through nature inspired computing for efficient glaucoma classification from fundus images," *Multimedia Tools and Applications*, pp. 1-72, Feb. 2024, doi: 10.1007/s11042-024-18624-y.
- [3] B. Al-Bander, B. M. Williams, W. Al-Nuaimy, M. A. Al-Tae, H. Pratt, and Y. Zheng, "Dense fully convolutional segmentation of the optic disc and cup in colour fundus for glaucoma diagnosis," *Symmetry (Basel)*, vol. 10, no. 4, Apr. 2018, doi: 10.3390/sym10040087.
- [4] C. De Vente *et al.*, "AIROGS: artificial intelligence for robust glaucoma screening challenge," *IEEE Transactions on Medical Imaging*, vol. 43, no. 1, pp. 542-557, Jan. 2024, doi: 10.1109/TMI.2023.3313786.
- [5] S. C. Patel and M. I. Patel, "Analysis of CDR of fundus images for glaucoma detection," in *Proceedings of the 2nd International Conference on Trends in Electronics and Informatics, ICOEI 2018*, pp. 1071-1074, Nov. 2018, doi: 10.1109/ICOEI.2018.8553707.

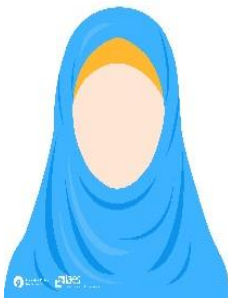
- [6] A. Issac, M. Parthasarathi and M. K. Dutta, "An adaptive threshold based algorithm for optic disc and cup segmentation in fundus images," in *2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)*, Noida, India, 2015, pp. 143-147, doi: 10.1109/SPIN.2015.7095384.
- [7] A. I. Saifi, P. Nagrale, K. K. Ansari, I. Saifi, and S. Chaurasia, "Advancement in Understanding Glaucoma: A Comprehensive Review," *Cureus*, vol. 15, no. 9, Sep. 2023, doi: 10.7759/CUREUS.46254.
- [8] L. K. Singh, M. Khanna, S. Thawkar, and R. Singh, "Collaboration of features optimization techniques for the effective diagnosis of glaucoma in retinal fundus images," *Advances in Engineering Software*, vol. 173, p. 103283, Nov. 2022, doi: 10.1016/J.ADVENGSOFT.2022.103283.
- [9] H. A. Nugroho, W. K. Z. Oktoeberza, A. Erasari, A. Utami, and C. Cahyono, "Segmentation of optic disc and optic cup in colour fundus images based on morphological reconstruction," in *2017 9th International Conference on Information Technology and Electrical Engineering, ICITEE 2017*, Jul. 2017, pp. 1–5, doi: 10.1109/ICITEED.2017.8250456.
- [10] N. A. Zainudin, A. Nazari, M. M. Mustafa, W. N. S. W. Zakaria, N. S. Suriani, and W. N. H. W. Kairuddin, "Glaucoma detection of retinal images based on boundary segmentation," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 1, pp. 377–384, Apr. 2020, doi: 10.11591/ijeecs.v18.i1.pp377-384.
- [11] A. Septiari, H. Hamdani, and D. M. Khairina, "The contour extraction of cup in fundus images for glaucoma detection," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 6, no. 6, pp. 2797–2804, Dec. 2016, doi: 10.11591/ijece.v6i6.pp2797-2804.
- [12] L. K. Singh, M. Khanna, H. Garg, and R. Singh, "Efficient feature selection based novel clinical decision support system for glaucoma prediction from retinal fundus images," *Medical Engineering & Physics*, vol. 123, p. 104077, Jan. 2024, doi: 10.1016/J.MEDENGPY.2023.104077.
- [13] L. K. Singh, M. Khanna, H. Garg, and R. Singh, "Emperor penguin optimization algorithm- and bacterial foraging optimization algorithm-based novel feature selection approach for glaucoma classification from fundus images," *Application of Soft Computing*, vol. 28, no. 3, pp. 2431–2467, Feb. 2024, doi: 10.1007/s00500-023-08449-6.
- [14] J. Civit-Masot, M. J. Dominguez-Morales, S. Vicente-Diaz, and A. Civit, "Dual machine-learning system to aid glaucoma diagnosis using disc and cup feature extraction," *IEEE Access*, vol. 8, pp. 127519–127529, 2020, doi: 10.1109/ACCESS.2020.3008539.
- [15] J. Civit-Masot, F. Luna-Perejon, S. Vicente-Diaz, J. M. R. Corral, and A. Civit, "TPU cloud-based generalized U-Net for eye fundus image segmentation," *IEEE Access*, vol. 7, pp. 142379–142387, 2019, doi: 10.1109/ACCESS.2019.2944692.
- [16] A. Sevastopolsky, "Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network," *Pattern Recognition and Image Analysis*, vol. 27, pp. 618–624, Apr. 2017, doi: 10.1134/S1054661817030269.
- [17] J. G. Zilly, J. M. Buhmann, and D. Mahapatra, "Boosting convolutional filters with entropy sampling for optic cup and disc image segmentation from fundus images," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9352, pp. 136–143, 2015, doi: 10.1007/978-3-319-24888-2_17.
- [18] L. K. Singh, M. Khanna, S. Thawkar, and R. Singh, "Deep-learning based system for effective and automatic blood vessel segmentation from Retinal fundus images," *Multimedia Tools and Applications*, vol. 83, no. 2, pp. 6005–6049, Jan. 2024, doi: 10.1007/S11042-023-15348-3.
- [19] O. J. Afolabi, G. P. Mabuza-Hocquet, F. V. Nelwamondo, and B. S. Paul, "The use of U-Net lite and extreme gradient boost (XGB) for glaucoma detection," *IEEE Access*, vol. 9, pp. 47411–47424, 2021, doi: 10.1109/ACCESS.2021.3068204.
- [20] P. V. Ramesh *et al.*, "Utilizing human intelligence in artificial intelligence for detecting glaucomatous fundus images using human-in-the-loop machine learning," *Indian Journal of Ophthalmology*, vol. 70, no. 4, pp. 1131–1138, Apr. 2022, doi: 10.4103/IJO.IJO_2583_21.
- [21] M. Anusha, S. Devadharshini, F. M. Farook, and G. Ananthi, "Glaucoma detection using the YOLO V5 algorithm," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 202–212, 2023, doi: 10.1007/978-3-031-44084-7_20.
- [22] F. Fumero, S. Alayon, J. L. Sanchez, J. Sigut, and M. Gonzalez-Hernandez, "RIM-ONE: an open retinal image database for optic nerve evaluation," *2011 24th International Symposium on Computer-Based Medical Systems (CBMS)*, Bristol, UK, 2011, pp. 1-6, doi: 10.1109/CBMS.2011.5999143.
- [23] J. Sivaswamy, S. R. Krishnadas, G. D. Joshi, M. J. Ujjwal, and S. Tabish, "Drishti-GS: retinal image dataset for optic nerve head (ONH) segmentation," in *2014 IEEE 11th International Symposium on Biomedical Imaging*, 2014, pp. 53-56, doi: 10.1109/ISBI.2014.6867807.
- [24] E. J. Carmona, M. Rincón, J. García-Feijoó, and J. M. Martínez-de-la-Casa, "Identification of the optic nerve head with genetic algorithms," *Artificial Intelligence in Medicine*, vol. 43, no. 3, pp. 243–259, Jul. 2008, doi: 10.1016/J.ARTMED.2008.04.005.
- [25] Z. Zhang *et al.*, "ORIGA(-light): an online retinal fundus image database for glaucoma analysis and research," *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, Buenos Aires, Argentina, 2010, pp. 3065-3068, doi: 10.1109/IEMBS.2010.5626137.
- [26] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT: real-time instance segmentation," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2019, pp. 9156–9165, Oct. 2019, doi: 10.1109/ICCV.2019.00925.
- [27] S. Saha, J. Vignarajan, and S. Frost, "A fast and fully automated system for glaucoma detection using color fundus photographs," *Scientific Reports*, vol. 13, no. 1, pp. 1–11, Oct. 2023, doi: 10.1038/s41598-023-44473-0.
- [28] A. H. Ashraf *et al.*, "Weapons detection for security and video surveillance using CNN and YOLO-V5s," *Computers, Materials & Continua*, vol. 70, no. 2, pp. 2761–2775, Sep. 2021, doi: 10.32604/CMC.2022.018785.
- [29] E. Elhariri, N. El-Bendary, and S. M. Saleh, "Strawberry-DS: Dataset of annotated strawberry fruits images with various developmental stages," *Data Brief*, vol. 48, p. 109165, Jun. 2023, doi: 10.1016/J.DIB.2023.109165.
- [30] S. K. Shandilya, A. Srivastav, K. Yemets, A. Datta, and A. K. Nagar, "YOLO-based segmented dataset for drone vs. bird detection for deep and machine learning algorithms," *Data Brief*, vol. 50, p. 109355, Oct. 2023, doi: 10.1016/J.DIB.2023.109355.
- [31] R. Xu, H. Lin, K. Lu, L. Cao, and Y. Liu, "A forest fire detection system based on ensemble learning," *Forests*, vol. 12, no. 2, pp. 1–17, Feb. 2021, doi: 10.3390/f12020217.
- [32] O. Kovalyk, J. Morales-Sánchez, R. Verdú-Monedero, I. Sellés-Navarro, A. Palazón-Cabanes, and J. L. Sancho-Gómez, "PAPILA: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment," *Scientific Data*, vol. 9, no. 1, pp. 1–12, Jun. 2022, doi: 10.1038/s41597-022-01388-1.
- [33] A. Sánchez-Morales, J. Morales-Sánchez, O. Kovalyk, R. Verdú-Monedero, and J. L. Sancho-Gómez, "Improving glaucoma diagnosis assembling deep networks and voting schemes," *Diagnostics (Basel)*, vol. 12, no. 6, Jun. 2022, doi: 10.3390/DIAGNOSTICS12061382.





BIOGRAPHIES OF AUTHORS

Mai Hesham     received the engineer degree in biomedical engineering from Higher Technological Institute (HTI) in 2019. Her research interests include, image processing, and programming. She can be contacted at email: maiheshamhti@gmail.com.



Ghada Kareem     received her B.Sc. degree in 1998 in Electronics and Communication Department from Zagazig University. In 2007, she received the M.Sc. degree in control and computer system from Zagazig University, in 2016 she received the Ph.D. degree in control and computer system from Zagazig University. She is currently and since July 2016 lecturer in Higher Technological Institute in Cairo, Egypt. Her research interests include adaptive control, image processing, signal processing, and sliding control. She can be contacted at email: ghada.gameel@hti.edu.eg.



Marwa Hadhoud     is an Associate Professor at the Department of Biomedical Engineering, Faculty of Engineering Helwan University. She received her B.Sc., M.Sc. in the biomedical engineering from Faculty of Engineering Helwan University in 2003 and 2007 respectively. In 2012 she received a dual degree Ph.D. in biomedical engineering from Faculty of Engineering Helwan University and Politecnico di Torino Italy. Her research interests are medical signal/image processing, pattern recognition, and bioinformatics. She can be contacted at email: marwa_hadhoud@h-eng.helwan.edu.eg.